Secondary Services and the Rising Tide of Paper

J. MARTYN

SOME YEARS before the phrase "information explosion" became current, the author of the Book of Ecclesiastes observed that "Of making many books there is no end" and, voicing a view that one suspects is shared by many of those for whom information services are hopefully designed, added that "Much learning is a weariness of the flesh." Since that time, the journal literature has appeared, and the potential weariness of the flesh has been correspondingly increased. Strictly speaking, Ecclesiastes has fallen into the common trap of equating books with knowledge because today one still speaks of an information explosion when what is properly meant is a paper explosion; that is to say, there is a tendency to assume that because the number of published items is continually increasing, the fund of human knowledge is being added to by a similar amount. The distinction between information and documents is an important one, because what abstracting and indexing services are basically concerned with are pieces of paper, and the fact that there are a great many of them causes most of the problems of secondary services.

At various times the number of papers published annually has been estimated at anything up to a million and a half. Variations in estimates arise because of differences in the bases examined, in the definition of what constitutes a published paper, and in the selection of the area of knowledge examined. Most estimates concern themselves with science and technology, and in this field the estimate published by Vickery in 1968 is probably the most firmly based.[1] He defined a paper as an authored contribution to a scientific or technical journal, and, by careful sampling of current issues of journals held by the National Lending Library, arrived at an estimate of approximately 850,000 papers pub-

J. Martyn is a Research Scientist at the Aslib Department of Research and Development, London.

lished annually. Comparing this figure with other estimates, it seems reasonable to suppose that by the end of the 1960s the annual output of papers in science and technology was of the order of 1 million.

A more recent study by Vickery[2] of the growth of journal literature, from its first appearance to the present, suggests an overall approximate doubling period of twenty years, and a cumulative total of 30 million papers. On these estimates as foundation one can rear an imposing edifice of entertaining speculation. For example, the British Museum having between seventy and eighty miles of shelving, it would be possible to house one copy of every scientific and technical paper ever published on the existing shelves, if they were empty. As there would be some room to spare, it follows that the volume of books at present housed there is greater than the volume of papers that currently exists. Or, to put it another way, there are more words in books than there are in journals. Considering the rate at which the British Museum adds new shelving, it is likely that the doubling period of the book literature is about twenty-three years, not grossly dissimilar to that of the journal literature. It follows that not only are there more words in books than in journals, but there always will be, and the gap between the two is widening.

As it is clearly harder to read all the books than to read all the journal papers, why not abstract books, rather than papers? Of course, the foregoing flight of fancy has neglected the journal literature of social science and the humanities, as one often does, and has glossed over the proportion of the British Museum library that is fiction, social science or humanities, or is simple translation of other volumes, and tends also to ignore the higher repetition rate of information in books as compared with journal papers. Nevertheless, it was an example of the sort of extrapolation one can (and frequently does) engage in in the information field, when presented with a little nugget of near-fact.

Vickery's estimate of 30 million papers referred to science, technology and medicine only. It is difficult to extend the estimate to the social sciences and the humanities, partly because of the lack of comprehensive listing, partly because of difficulties in defining social science (or humanity), and partly because the more literary a document becomes, the less likely is it to be classed by an information scientist as literature; that is to say, the scientific paper is an inherently more definable unit act of communication in science than the vehicles of communication appropriate to other scholarly activities.

There are a great many secondary services whose purpose it is to

provide references to, and frequently abstracts of, portions of the total mass of literature, often in an indexed or classified form. The portions chosen can be either topic or discipline oriented like *Chemical Abstracts,* mission oriented like *Food Science and Technology Abstracts,* quality oriented like the *Science Citation Index,* or with a national orientation like *British Technology Index.* The coverage policy can be selective, usually on a quality basis, or comprehensive. The total number of secondary services is uncertain. In 1963 the National Federation of Science Abstracting and Indexing Services published a listing of all the science-technology services publishing more than 100 abstracts a year that they had been able to identify, and this listing contained 1,855 entries;[3] NFSAIS is currently working on a revision of this list. Separate sections of large services, such as the parts of *Referativny Zhurnal,* were shown as individual entries, and a number of the services listed were abstracts sections of otherwise normal journals. Some services were available in several formats, e.g., printed codex or cards. There is no doubt that the forthcoming revised list will show considerable variation from the 1963 list, particularly in terms of the variety of formats available, since there are now in the region of 100 services available on magnetic tape.

In the intervening period there has also been an accelerated tendency for large services to subdivide or to provide varieties of bibliographic output, as for example *Chemical Abstracts Condensates, Chemical Titles, Chemical-Biological Activities,* and so on. And undoubtedly the numbers of available services have increased. Each individual service will also certainly show an increase in the numbers of items published annually. *Chemical Abstracts* has jumped from 269,293 in 1967 to 350,105 in 1971, and in the same period INSPEC has jumped from 71,032 to 148,944 and *Engineering Index* from 51,670 to 85,243. Other services will have increased similarly, partly as a response to increases in the primary literature, but also because of improved performance in acquiring the relevant literature, and especially because of policy decisions whose effect is to widen the scope of the secondary services by including much material which might previously have been regarded as peripheral.

Vickery, in the growth study referred to above,[2] also points out that an examination of the growth of six major abstracting services over the past decade shows a doubling period of six and a half years, which is a faster rate of increase than that of the primary literature itself. A recent study of information on tape,[4] carried out by the Aslib Research and

Development Department, identified about seventy services for which it was possible to estimate the total number of entries made up to December 1970, and the annual number of entries. This study indicated that at least 3.5 million machine-readable entries are now recorded annually, and the total number of entries on tape was, at the end of 1972, about 24 million.

It is already obvious that there are many more entries made annually in secondary services than there are primary papers published. Since no service is notifying imaginary references, it follows that there is very considerable overlap among the secondary services, with multiple notification of many items. It is worth noting that the great majority of services examined in the Aslib tape study are in the English language, so there is no evidence to support the hypothesis that the overlap is caused by abstracting or notification in different languages: it would not be unreasonable to assume that, had it been possible to include non-English language services in the study, a higher rate of duplicate notification would have been suggested.

Martyn[5] noted that the study of a number of reasonably comprehensive specialized bibliographies showed that about one-fifth of the references contained in the bibliographies had not been picked up by any major secondary service, and concluded, with the specialized bibliographies used being drawn from a number of areas of science and technology, that it was not unfair to assume that, over the whole field, a similar situation prevailed. A more recent unpublished work by Martyn, in which items notified by the Biodeterioration Information Centre have been sought in Institute for Scientific Information coverage, *Chemical Abstracts* services, *Biological Abstracts, Index Medicus* and *Food Science and Technology Abstracts* (all services which may be searched by computer), has confirmed the original view because, of the 1,874 references to 1969 journal literature sought, more than 23 percent proved not to have been covered by the services examined.

We now have the picture that, so far as science and technology are concerned, roughly 1 million items of primary journal literature appear each year. Probably (augmenting the Aslib tape coverage figure to allow for services not yet on tape) more than 4 million entries appear annually in secondary services. As approximately one-fifth of the items published are not notified, it is reasonable to conclude that, of those items which *are* notified, each is notified, on the average, five times. It may further be assumed that many of the notified items are notified many more than five times. This is a result of the natural evolution of

the system of coverage of the literature of science and technology. The major services, which are mainly the longest established, cover the literatures of specific disciplines; coverage of their core literatures is generally complete, or nearly so, but coverage tends to fall away from completeness in the fringe areas, as is natural. The core of a discipline is clearly defined, and the relevance of a document to that core can be easily assessed, but a fringe area is by its nature less clearly defined, and decisions of relevance become more subjective. Usually a major service is operated by full-time professionals, with paid abstractors and indexers, and a marketing system organized along commercial lines. Exchange of appropriate material among services has been common for a long time, and is gradually becoming more systematic.

Together with the major services there is a population of smaller secondary services which cover subdisciplines, whose coverage is often wholly contained within the major services, but which continue to exist either because they provide their customers with more specific information (e.g., the user has to buy less of what he does not need in order to get what he does need), or because the quality of their product is superior by virtue of being better indexed or having more information-content per item notified. These services also tend to be reasonably long-standing and professionally organized, but not infrequently rely on voluntary or semivoluntary (i.e., underpaid) assistance with their input.

There are also a large number of small services which have come into being in order to cover interdisciplinary areas, the areas which lie between the defined coverage areas of the bigger services or in some cases outside them. Being interdisciplinary, they tend to serve developing areas of science which are relatively novel, and, therefore, often of no great age. Because the areas of interest are novel, there is often not a very large population of users to support them, so their potential income is small and they rely on volunteer labor, and the absorption of their overheads by large organizations such as, for example, universities which provide accommodation and facilities, or major libraries which allow free access to their journal stocks, to keep their apparent costs down. Some are run by professional information workers and some are not. In many cases the marketing of their products is far from perfect. Where diversity can exist, it does.

Some services attempt to cover their chosen area of interest comprehensively and include everything relevant which can be discovered by scanning a wide range of journals, like *Chemical Abstracts* which "in

1971 . . . combed the contents of some 12,000 scientific and technical journals published in 56 different languages and chemical patents issued by 26 nations." Some restrict their coverage to journals alone, or patents or the report literature. Some cover selectively, imposing a degree of quality control on the material accepted for inclusion. Some simply claim to cover a field as well as they are able by drawing from a defined set of sources, like the *Bibliography of Reproduction* which claimed an approximate 50 percent coverage of the field of mammalian reproduction. ISI, unlike the majority, covers a stated list of journals comprehensively, including everything published in the covered journals. The quantity of information presented varies from a bare reference to a detailed informative abstract, often acceptable as a substitute for the original document. The style and quality of indexing varies similarly, and the time elapsed between publication of a document and its notification can vary from two or three weeks to, in extreme cases, several years.

Just as the primary literature grows, develops, overlaps, multiplies and occasionally dies with little to check it but the pressures of the market (moderated sometimes by direct subsidies, or the unrealized subsidies that arise from defective accounting), so the secondary literature exhibits the richness, the vigor of growth, the variety, the extravagance, the untidiness and the illogicality characteristic of the majority of human activities in a free society. Irrationality is the price of freedom in more than a purely philosophical sense, and even in those systems where sanctions can be imposed to enforce order, humanity, like cheerfulness, keeps breaking through, and the systems therefore fall away from perfection.

What the effects of this lack of systems are on the user depends very much on the nature of the user. If his interests lie within the boundaries of a long-established scientific discipline, he is liable to find the present system adequate to his needs, particularly if his requirement is for less than total coverage of his topic of interest. As total coverage is liable to produce references to many more documents than he is willing or able to read, he is predisposed to accept partial coverage, although welcoming some assurance that what he misses is of less importance than what he finds. If, however, as is perhaps more commonly the case, his immediate requirement is for a piece of information whose normal habitat is in a discipline outside his own, or he requires a body of references which are scattered among the literatures of several disciplines, or which are even difficult to locate within the normal framework of

the traditional disciplines—in other words, if his temporary or enduring need is for interdisciplinary information—he faces serious problems. The first problem is the choice of service to search, which is largely a matter of trial and error. Having selected a service, if he is seeking a reasonably comprehensive or, better, representative, sample of literature on his interdisciplinary topic, he can be certain that all his needs will not be met from within that one service alone. He must therefore turn to other services, but if he does this, he can be certain that not only will he be unlikely ever to achieve anything approaching comprehensive coverage, but also he will inevitably find the same items over and over again.

The first problem may or may not be serious. There are some grounds for supposing that the information that is not picked up by the secondary services tends to be of a lower level than that which is picked up; lower in the sense of generally not being the first appearance of the contained information, but rather rewritten versions of it produced for scientists in different fields or technologists in the same field as the field of origin of the original information to which it refers.

The second problem is more important, particularly in the case of an individual searching services based on machine-readable records, because in the majority of cases the payment for records retrieved from these services is directly related to the number of items retrieved or expected to be retrieved, and while the user population may become resigned to the prospect of paying for a number of unwanted references inevitably retrieved with those which are wanted, it is less likely to accept the necessity of paying afresh for items which it has already bought, however relevant they may be.

What is the purpose of the system of secondary services? It is seldom stated in a global context, being usually related to a particular population. Any specific secondary service is generally aimed at covering the literature which is likely to be of use or interest to scientists or others in a stated discipline or subdiscipline, or with a stated mission; it tends, that is to say, to be user-oriented. This unfortunately means that while there are very many users whose average needs are satisfactorily met by existing single services, there are also, inevitably, many other users, either in interdisciplinary areas or in some areas of applied science and technology, whose needs are not met, just as a normally well-served user occasionally has an unusual information need which is difficult to meet from his habitual sources. In an ideal world, access to all parts of the primary literature should be equally possible for all potential users,

and to afford such access should be the object of the system of secondary services as a whole.

To insure equal access to all documents, it is necessary to insure that references to all documents exist somewhere in the overall system. So far as the book literature is concerned, this is gradually being achieved through the agency of the various national bibliographies, and when all countries are eventually producing MARC-like services for their own book output, then, errors and omissions excepted, world coverage of monograph material will be complete. It should be noted that the existence of national bibliographies does not preclude the existence of specialized bibliographies, but rather facilitates their production.

National bibliographies work because they are organized on the basis of the physical properties of the items they cover; they are identifiable as books, according to an agreed definition, and they are published in identified locations. Discipline- or mission-oriented secondary services are less successful in terms of coverage because they are based on the intellectual properties of the documents they cover—what they are about, rather than where they come from. A "journal-paper MARC" operated on a national basis is as feasible as existing MARC systems, and would at least provide comprehensive coverage, searchable by journal title, article title and author, of the journal literature of the individual countries operating such a system.

Such a system would take some time to implement, and might well require legislation, if only to provide enabling funds. It would inevitably be a considerable time before the world's literature could be covered in this way. Attempts to approximate a consolidated coverage of the journal literature could be made if it were practicable to merge the existing major tape-based secondary services in such a way as to create a unified file containing the author, title and reference of each item contained in the major services, and perhaps augment it with additional material known not to be covered. One problem which would have to be faced would be identification of noncovered material, and this would be greatly facilitated if the major services could state categorically which journals they cover completely. (This would be of value immediately, in that searchers of more than one service could use a logical NOT to prevent re-retrieval of at least some of the items found in the first search.)

Another problem would be the elimination of duplicate references, a problem which is currently intensified by the great variation in timelag of available systems. A reference appearing in service A perhaps a

month after first publication may appear in service B after a further three months, and again in service C six months later still; this problem can be met by using a journal-item-received register file and automatically deleting from each fresh input tape every item already shown to be held, but it is a cumbrous and unwelcome necessary addition to a notional system. The result, however, would be a single source which could be searched by a user without incurring cash penalities for duplicate retrieval.

There are basically three routes towards a unified, controlled system of secondary services. They are to operate the system yourself, with or without using the elements already existing, to coerce others into operating the system, or to persuade them to do so. The first route is essentially what has been suggested here. Coercion is not only morally undesirable, but also not practicable, so may be dismissed. Persuading others, usually known as "international cooperation," is the route which has been and is most often followed, not without a certain amount of success, but anyone with any experience of the international information field knows this is an arduous route, bedeviled with questions of national, organizational and individual self-interest. One recognizes that in the information field, as in other perhaps more important or more urgent problem areas, international cooperation is the only feasible long-term solution, but it is suggested that until this happy state be reached, some stop-gap operations would not be unuseful to the actual users of information, for whom, after all, the secondary services are notionally intended, and who, ultimately, pay the bill.

References

1. Vickery, B. C. "Statistics of Scientific and Technical Articles," *Journal of Documentation,* 24:192-95, Sept. 1968.

2. ———. *The Growth of Scientific Literature.* (To be published)

3. National Federation of Science Abstracting and Indexing Services. *A Guide to the World's Abstracting and Indexing Services in Science and Technology* (Report no. 102). Washington, D.C., NFSAIS, 1963.

4. "Information on Tape." Aslib Research & Development Department, Aslib, 1971. (Internal report)

5. Martyn, John. "Tests on Abstracts Journals: Coverage, Overlap and Indexing," *Journal of Documentation,* 23:45-70, March 1967.